VAA-Weekly-Progress

09/05 - 09/10

VIP

PURDUE UNIVERSITY

Context

- Last week, we reviewed some papers on semi-supervised keypoint localization and the visual transformer pose estimator
- We decided for this week to focus on solidifying a baseline model to proceed with, and comparing models pretrained on different datasets

Deciding on a Baseline System

- OpenMMLab?
- HRNet vs. RTMPose?
  - Read?
  - Performance?
    - Established benchmark datasets (e.g., AP-10k, AnimalPose)
    - Try your own images
      - What may cause trouble for pose estimators?
- Pre-trained on AP-10k or pre-trained AnimalPose?
  - What are the benefits and drawbacks?
  - Why not both?
  - What does the dataset evaluation system look like?
    - How did AP-10k tackle the detection problem?

Deciding on a Baseline System

- What do you choose?
- Why do you choose it?
- What is needed to get it working?
- What could go wrong?

HRNET Evaluation

Trained on Animal Pose (w32 config):

Pros:

- More accurate
  - Placement of wither kp on body focused pictures (on a higher place)
- Marks both neck and wither
- More accurate kp placement for average size blurry images

Cons:

- Excessive number of kp's on partially occluded objects
- Excessive/lack of kp's on weird poses
- Misplaces head kp's in blurry night time pictures

Trained on AP-10k (w48 config)

Pros:

- More accurate kp's for
  - Night time images (black and white)
  - Weird poses
  - Tracks hip kp more accurately in partial images
  - Small blurry images

Cons:

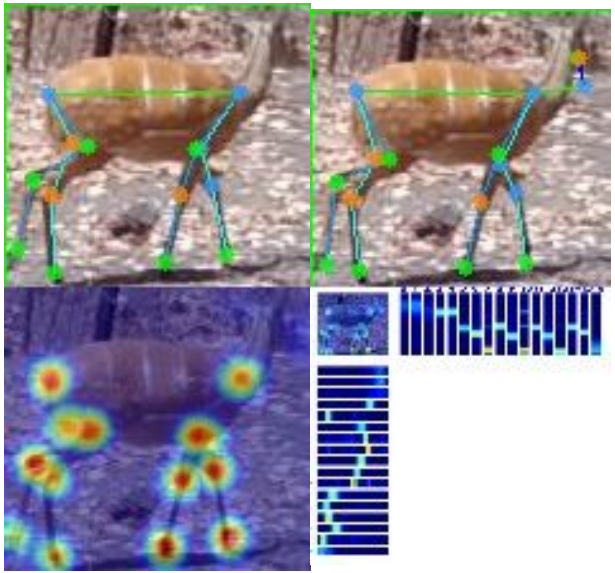
- Does not see the head in blurry night time images
- Adds additional hip kp in night time images

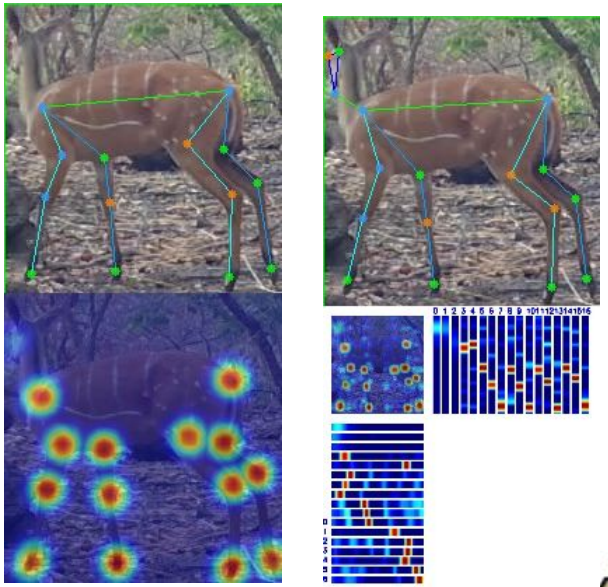
VIP

PURDUE UNIVERSITY

HRNet(w48) vs RTMPose

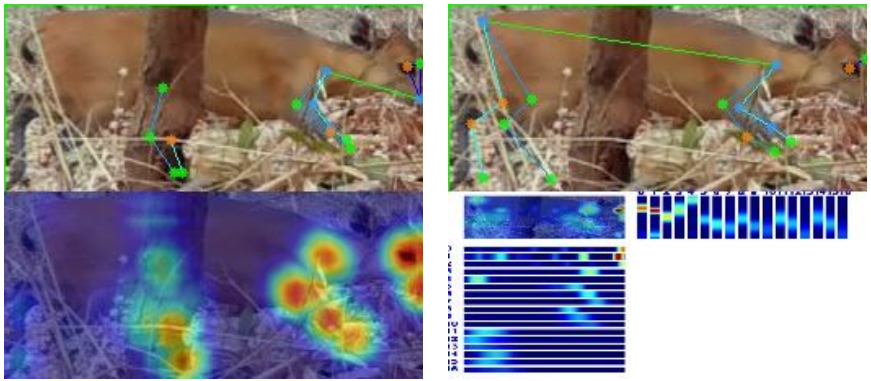
We ran both HRNet and RTMPose (pretrained on AP-10k) on our sample of antelope data. It appears that when predicting keypoints on the same image, RTMPose does a better job on detecting keypoints on the head.

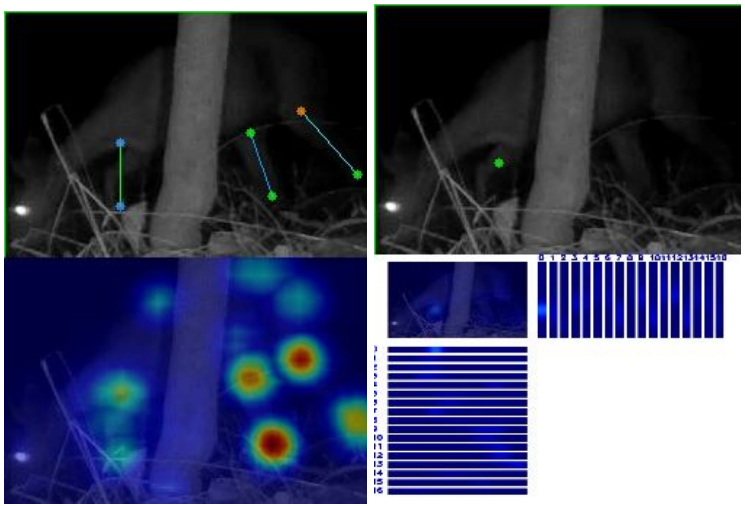




HRNet(w48) vs RTMPose

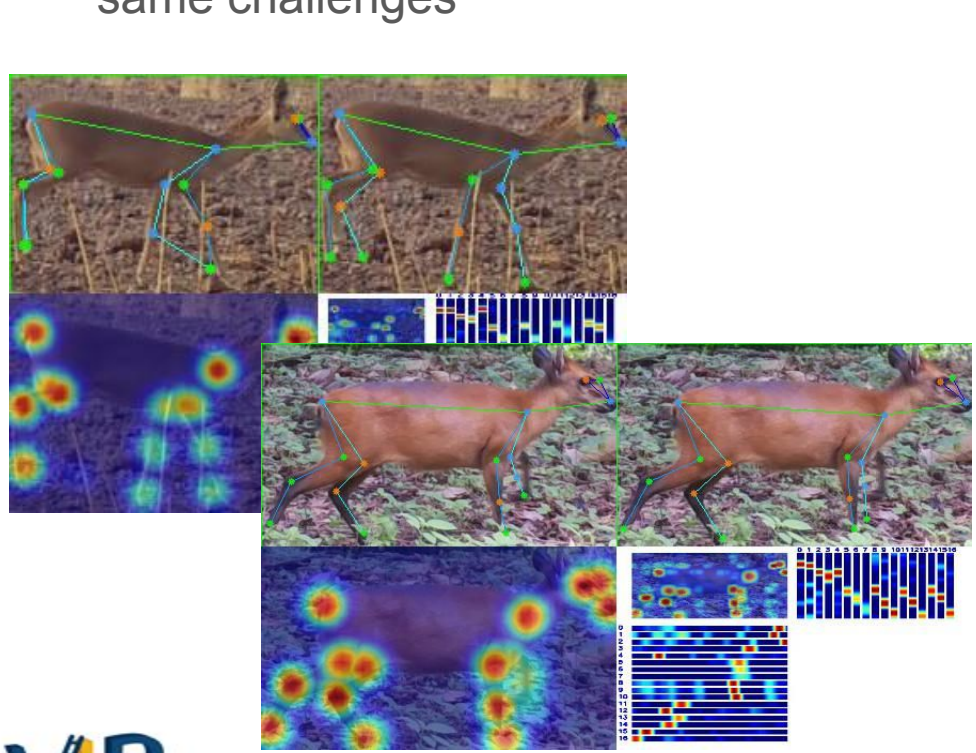
When detecting keypoints on an image with some obstruction, both models have difficulty. Note: in other night-time images, the two models perform at similar levels.

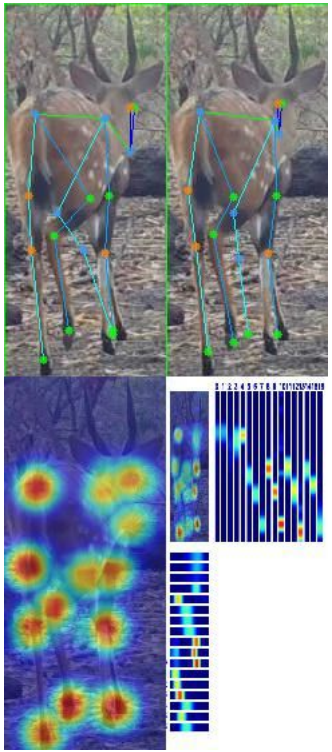


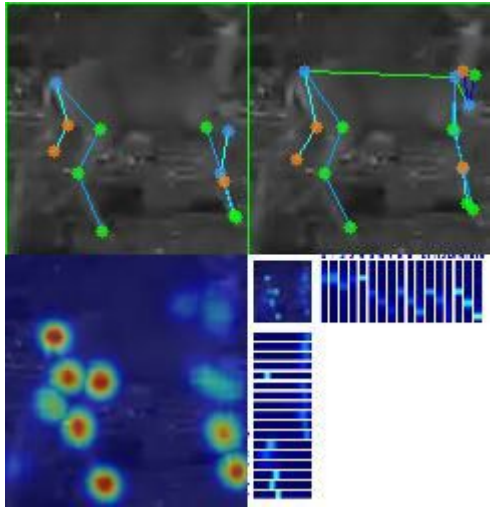


HRNet(w48) vs RTMPose

RTMPose seems to be slightly more robust, but both models suffer from the same challenges







RTMPose Evaluation

Trained on AP-10k:

Pros:

- Tends to work well for when detecting the head using kps
- Cut off images are labeled fairly well
- Twisted/Weird poses still have relatively accurate kps
- Some night time images have ideal kps

Cons:

- Struggles with obstacles blocking animal of interest
- Some blurry night images lack kps in necessary areas

Final Thoughts

We decided to proceed with RTMPose (trained on AP-10K) as our baseline model, because based on a qualitative analysis with a small subset of our data, it appears to have better performance compared to HRNet, specifically in images with occlusion and overall does a better job of identifying keypoints.

Next Steps

- Train RTMPose on Animal Pose data set and compare and contrast qualitative performance to determine on which training set results in better qualitative performance of RTMPose
- Combine Animal Pose and AP-10K into one point(converting wither and neck points in Animal Pose to neck in AP-10K) and train RTMPose on the combined data set and test qualitative performance against RTMPose trained on each data set separately
- Consider possible use of animal data sets as training data sets(AP-36K)

Personal Progress

Medha

- Annotated 100 images with labels (Set 4)
- Familiarized with OpenMMPose and analyzed outputs of two different models (HRNet and RTMPose) on our antelope data
- Working on setup for OpenMMPose on personal deadcat
- Met with team to decide on the baseline model to proceed with

Parth

- Annotated 100 images with labels (Set 6)
- Ran and analyzed RTMPose and HRNet Models trained AP-10K dataset
- Tested the models on our data from Week 1
- Decided on baseline model with team

Aryan

- Annotated images from the set 0 with labels
- Ran both models on previous data for a comparative analysis

Josh

- Annotated 100 images with labels (Set 1)
- Read abstracts and introductions of HRNet and RTMPose papers and watched talks to become familiar with the two models architectures and general advantages and disadvantages
- Analyzed HRNet and RTMPose on set 1 of antelope images to qualitatively determine which network is more suited to be our baseline system
- Discussed with team about which model to choose as baseline

Claire

- Annotated 100 images with labels (Set 2)
- Familiarized with OpenMMPose, and analyzed and compared results of running sorted antelope images with HRNet and RTMPose
- Discussed with team about which model to choose as baseline

Armaan

- Annotated 100 images with labels (Set 5)
- Wrote script batch.py to run a specified model and configuration over a whole folder of antelope images at once.
- Analyzed HRNet trained on animal pose vs trained on AP-10k across different configurations (w32, w48, resnet101) on set 5 of antelope images to qualitatively determine which model (and configuration) is most suited to be our baseline system
- Discussed with team about which model to choose as baseline

Shaan

- Labeled 100 images
- Re-used some of Armaan’s code and put together a script to align outputs from pretrained HRNet and RTMPose models on our own antelope data
    - This streamlined our goal of conducting qualitative analysis as it allowed for viewing outputs on the same image side by side.